INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 3, March 2026 | ISSN: 2456-4184 | IJNRD.ORG

INNOVATIVE FRAMEWORK FOR ENHANCING CYBER SECURITY THROUGH BIG DATA ANALYTICS USING ML

[1]Pranali R. Landge
Research scholor
P.G. Department of Computer Science
Sant Gadge Baba Amravati University, Amravati
Maharashtra, India

[2]Dr. Swati S. Sherekar
Professor
P.G. Department of Computer Science
Sant Gadge Baba Amravati University, Amravati
Maharashtra, India

Abstract
The frequency and complication of cyberattacks have increased in an increasingly connected world, presenting serious risks to cyber security. The sheer volume and speed of data created by contemporary networks are constantly too important for traditional security mechanisms to handle. This study investigates how big data analytics might meliorate cyber security by using massive databases to find trends, spot irregularities, and anticipate possible risks. We suggest a system for assaying network business and security records that combines big data technologies with slice- edge ML styles. Big data analytics may greatly meliorate cyber security postures by furnishing a visionary line of protection against changing online risks. By offering a thorough fashion for incorporating big data into security fabrics and pointing out motifs for further exploration, this study adds to the expanding corpus of knowledge in cyber security.
Keywords Big data analytics, ML, cyber security

1. Introduction

Because cyber risks are constantly changing and getting more sophisticated, cyber security is a constant issue for businesses in all sectors. A 2023 analysis by Cybersecurity gambles systems that the periodic cost of cybercrime would increase from$ 3 trillion in 2015 to$ 10.5 trillion by 2025. The rise in cybercrime emphasizes how important it's to have strong and flexible security measures. By making monitoring, anomaly discovery, and predictive trouble analysis possible, big data analytics provides a revolutionary approach to cyber security. 65 of businesses are administering big data technology to meliorate their cyber security systems, according to a McKinsey [2022] study. According to disquisition, mortal mistake or a failure to recognize vulnerabilities accounts for 80 of cyber security breaches, pressing the necessity of sophisticated logical tools. Organizations can transition from reactive to visionary security operation by using big data analytics, which can exercise vast amounts of eclectic data and find patterns that would be impossible to find by hand. By automating trouble discovery and response, machine knowledge and big data analytics have the eventuality to cut down on the average time it takes to identify and neutralize risks by over to 70 when compared to farther conventional approaches.

Table 1: Inspiring Data to Enhance Cybersecurity with Big Data Analytics

SN	Statistics	Value/Score
1	Projected global cybercrime cost [2025]	$10.5 trillion annually [Cyber security Ventures, 2023]
2	Cyber security breach due to human error	80% [IBM, 2021]
3	Percentage of companies adopting Big Data for cyber security	65% [McKinsey & Company, 2022]
4	Average time to detect and mitigate threats with traditional methods	6-9 months [Verizon, 2022]
5	Reduction in threat detection time with Big Data and AI	70% [Gartner, 2022]
6	Organizations experiencing a cyber-attack annually	43% [Ponemon Institute, 2023]

These figures illuminate how important it's to use slice- edge technologies like big data analytics to address the raising cyber security issues. By using big data, associations can enhance their security posture, reduce mortal error, and wharf response times to implicit risks. With the adding dependence on digital platforms for communication, banking, shopping, and social commerce, cybercriminals have set up more openings to exploit vulnerabilities in technology systems. The bar graph presented below demonstrates the growth in the number of reported cybercrime cases in India from 2020 to 2024. Cyber Crime not only poses a trouble to the security of individualities and associations, but it also has significant profitable impacts. As digital transformation accelerates, so do the financial losses beget by cybercriminals conditioning.

The graph above illustrates the substantial increase in the financial losses due to cybercrimes in India from 2020 to 2024 Cybercrime has come one of the most significant risks to the global economy, with financial losses soaring over the formerly numerous times. As

technology continues to evolve, cybercriminals are getting increasingly sophisticated, targeting individualities, businesses, governments, and critical structure across the globe.

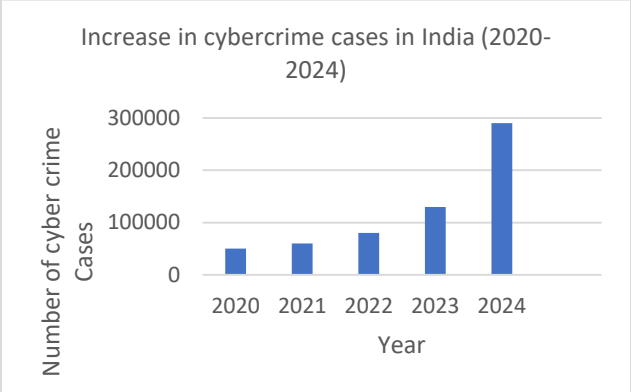

Fig 1: Increase in Cybercrime cases in India

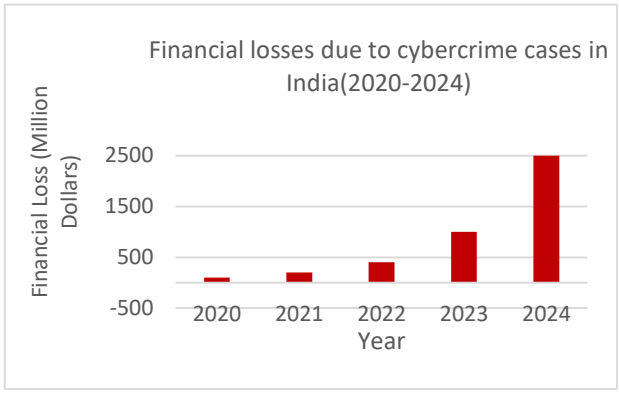

Fig 2: Financial losses due to Cybercrime in India

The graph below illustrates the dramatic rise in global financial losses caused by cybercrime from 2020 to 2024, reflecting the intimidating trend of digital risks worldwide. The data presented in the graphs on adding cybercrimes in India and encyclopedically from 2020 to 2024 highlights a concerning and accelerating trend. Cybercrime has evolved into a pervasive trouble, significantly impacting individualities, businesses, governments, and husbandry worldwide. The exponential rise in both the number of reported cybercrime cases and the financial losses reflects the growing complication and reach of cybercriminal exertion. Fig 3 Global financial Losses Due to cybercrimes In India, the swell in cybercrime cases and the associated financial losses demonstrate the vulnerability of a swiftly digitizing society. The increase in cybercrime from 2020 to 2024 emphasizes the need for enhanced cybersecurity measures, stricter enforcement of regulations, and wide public awareness to palliate these risks. With cybercrime cases anticipated to exceed 2 million by 2024 and financial losses raising, both individualities and associations must prioritize cybersecurity to guard particular data and financial means. Encyclopedically, the financial losses due to cybercrime are projected to surpass$ 4 trillion by 2024. The adding reliance on digital technologies, analogous as pall- commerce, and artificial intelligence, has created new openings for cybercriminals, while the obscurity handed by the internet complicates sweats to track and seize perpetrators.

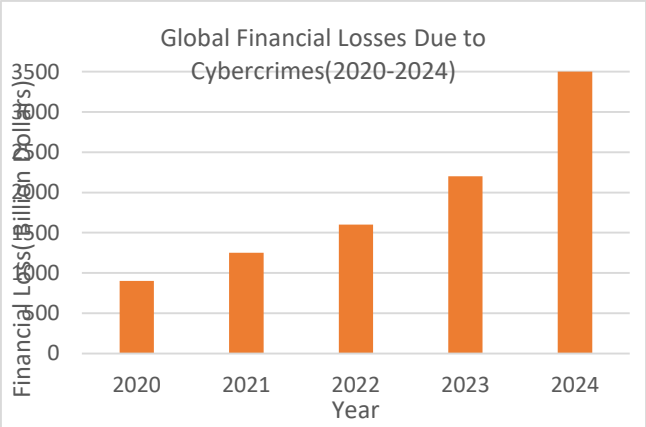

Fig 3: Global Financial Losses Due to cybercrimes

As the financial impact grows, it's apparent that both the public and private sectors must unite to strengthen cybersecurity fabrics, promote cyber hygiene, and foster global cooperation in combating cybercrime.

2. LITERATURE REVIEW

Big data technologies play a vital part in the process of collecting, assaying, and imaging massive and complex data. They help to discover retired patterns and to prize useful and sensitive information. Big data technologies are continuously supporting important network platform aiming to interconnect data realities and share information across them. still, big data is also an attractive zone for multitudinous security attacks that raise multitudinous challenges [1]. In fact, along with the high- speed growth of data that flows in networks, the network content scale is expanding, and the network terrain is more complicated. Farther cyber-attacks are getting decreasingly current and intricate. It's thus imperative to use dependable and robust Network Security Platforms [NSPs] that provides high- performance reports in the real and near real- time. nevertheless, Traditional NSPs [TNSPs], like any conventional tool, are shy to descry attacks in large and complex data within a reasonable time [2]. Effectively, Big data characteristics have a significant impact on NSPs. The once decade has seen rapid-fire progress in the internet, internet of effects and pall computing, and thereby a strong growth in data application for marketable and artificial operations. The term" Big Data" refers to massive data sets with large, different, and complex structures that are challenging to store, dissect, and handle with traditional processing approaches. For a big data handling approach to be effective, it should accommodate not only the size, haste, and diversity of similar data, but also their unique data protection conditions. nonetheless, there's a clear conflict between the pervasive operation of big data and affiliated security and

sequestration issues [1,2]. The only company with further than a million waiters encyclopedically is Google. Ten billion textbook dispatches are transmitted daily, and there are six billion mobile subscriptions worldwide. 50 billion widgets will be linked to networks and the internet by 2020 [3]. Big data analytics has surfaced as a transformative technology that offers important capabilities to enhance network security. By employing vast volumes of data generated from network exertion, stoner geste , and trouble intelligence, associations can gain precious perceptivity that grease visionary defense mechanisms. The operation of big data analytics enables monitoring, anomaly discovery, and prophetic analysis, empowering security brigades to identify and respond to pitfalls with unknown speed and delicacy [4]. Through an analysis of case studies and the perpetration of advanced logical tools, this exploration work will contribute to a deeper understanding of how big data can transfigure network security strategies, eventually fostering a more flexible digital terrain [3,4]. Within the coming decade, the number of information will increase by 50 times, still, the number of information technology specialists who keep up with all that data will increase by 1.5 times [5].

Big data is a new paradigm that applies to datasets that are too big for constantly used software tools to handle, acquire, and process in a reasonable quantum of time. These kinds of datasets are constantly unshaped, come from a variety of sources, have a high volume, and have a high haste of data flux and exodus. further crucially, big data must be Value and Veracity in order to be used in marketable decision- timber [6]. pall computing platforms, scalable storehouse systems, MapReduce, and largely resemblant processing databases are employed to handle huge data. A traditional field of study called distributed systems looks into a range of distributed computing operations and technologies, including pall computing and Chart Reduce. [7]. Massive growth in big data makes it grueling to acquire, form, store, manage, partake, dissect, and fantasize using standard database software tools. The size of the digital data world reached 2.72 zettabytes in 2012. By 2015, it's anticipated to have grown to roughly 8 zettabytes of data, doubling every two times [8]. By 2019, multimedia data is prognosticated to increase by 70 and have a significant impact on Internet backbone business [9]. Distributed systems exploration continues to produce new and creative results from academia and assiduity as a result of new paradigms and technology. Map Reduce, for case, is a distributed programming paradigm and related perpetration that's extensively used to grease distributed calculation over massive, pall- grounded datasets [10].

In order to reduce the quantum of time demanded to admit the information demanded to make opinions, its study focuses on the effective association and overall evaluation of numerous types of complicated information [11]. This study empirically examined the viability of civic public security evaluation modeling by combining the PSR fashion, fuzzy sense model, and entropy weight approach. An indicator system for assessing the fundamentals of public security was established using the PSR (Protective Security Requirements) approach [12]. To confirm the equity of this modeling, the importing assignment procedure employed the Entropy approach. This study examined the geographical and temporal dynamics in this region. The study problem and the integrated methodology were unique, in addition to being an interesting approach. The findings are interesting, and at the conclusion are some suggestions. We suggest a case study on a mobile data system to address the emigrations from druggies' mobile bias in order to demonstrate this strategy [13]. The core of contemporary business and wisdom is big data and its analysis. Online deals, emails, pictures, audios, prints, clicks, logs, bulletins, queries, health records, social media relations, scientific data, detectors, mobile phones, and their operations are some of the sources of these data [4,13]. Information filtering, miscellaneous database analysis, outfit operating conditions, alarm information processing and analysis, and traditional network security situation operating conditions are all areas in which it's used. Although assessments offer security directors a wealth of analysis data, the maturity of them calculate on system log analysis [14]. There are multitudinous issues, similar as a single data source, crummy performance, and assessments that calculate too heavily on network operation staff experience [15].

At present, the security situation assessment of network security is also a major change. From the initial passive security system construction of network security, detection, defense, attack. Security situation assessment is one of the active security system constructions. But the security situation of global network has still at the starting stage at home and abroad. But the relevant technical theory is not mature, so it is urgent to find an effective and accurate assessment method [16]. BM indicates that every day 2.5 exabytes of data created 90% of the data produced in last two years [17]. This research work analyzed both the temporal and spatial dynamics in this area. Not only was this approach interesting, but the research issue and combined method were original [18]. The findings are intriguing, and at the conclusion are some suggestions. We suggest a case study on a mobile data system to address the emissions from users' mobile devices in order to demonstrate this strategy [19]. Furthermore, the predictive power of security systems can be greatly increased by utilizing machine learning algorithms in big data analytics setting. These subject matter experts can accomplish this by discretizing the data into predetermined classes [20].

Cyber-threats have become highly sophisticated, and have increasingly targeted not only large enterprises but also small firms and individuals. Improving cybersecurity through Big Data analytics has proven to be a viable solution, as it helps detect and prevent cyber-attacks by analyzing data from multiple sources in real-time [21]. Big Data analytics can identify irregularities and possible threats more successfully than conventional techniques by combining diverse data from network logs, user activity, and threat intelligence feeds. For example, behavioral analytics can help identify deviations from normal user activity, thus revealing insider threats or compromised accounts [22]. The Human Face of Big Data accomplished at a global project by Rick Smolan, consists of data acquired from sensors and medical equipment in satellites and identified mobile devices. This massive data collection has highlighted combinations of key statistics, including 78.4 billion gigabytes [GB] of mobile data generated every month [2013], 48 hours of video and audio uploaded every minute, 72 million hackers per day, 12 IEEE uploaded articles, 26 billion gigabytes of stolen and uploaded data, 6 billion smartphone users, 30 billion RFID sensors, 2.5 10^18 bytes of everyday data, and 231 million active users on Twitter. 571 new websites are created every minute of the day [23]. By integrating data from multiple stakeholders and ensuring compliance with regulatory requirements, Big Data analytics can help organizations build a more secure and scalable resource to combat cyber threats. City sensors, and difficult-to-use heterogeneous environmental data that combine to analyze, reason and consume, particularly in compact, diverse settings like buildings, transportation, or water energy [24].

3. EXISTING METHODOLOGIES & ANALYSIS:

In the methodology for "Improving cyber–Security Through Big Data Analytics", the approach involves several critical steps, each contributing to the overall system's effectiveness. These procedures start with a basic analysis of existing approaches, and then move on to feature engineering, data preparation, model selection, and performance assessment. First, log data from network devices and associated systems is gathered. These records, which are utilized to reconstruct network behaviors and spot irregularities or questionable activity, comprise traffic logs, firewall logs, and packet-level observations. During this phase, intrusion detection systems [IDS] are widely used to provide alerts about known and undiscovered threats. In order to identify valuable qualities that are indicative of security events, we refine the raw data using feature engineering. These characteristics could include error rates, byte transfer sizes, request frequency, and connection duration. Finding trends or abnormalities that can be connected to particular attack types, such DDoS, botnet activity, or malware infections, is the goal. The preprocessing step involves transformation, normalization, and cleaning to guarantee data quality. By eliminating noise, adding missing values, and formatting data to transform redundant or heterogeneous inputs into a readable format, these actions are crucial for getting the data ready for efficient analysis. The use of big data analytics techniques forms the basis of security analytics. We classify or cluster network security threats using both supervised and unsupervised learning algorithms. Supervised learning methods such as DT, SVM, and RF enable the system to differentiate between normal traffic and various types of attacks. Unsupervised techniques are used to detect unknown or emerging threats by identifying anomalies or deviations from established patterns. Apache Kafka, Apache Spark Streaming, and Apache Flink tools enable the system to monitor network data continuously, detect and respond to security incidents as they occur. The capabilities of the system are essential for minimizing the response time and reducing the impact of cyber threats.

The system's performance is then assessed using a range of evaluation criteria, including the area under the ROC curve [AUC], recall, precision, and F1-score. In addition to ensuring that the system can accurately detect malicious activity without producing an excessive number of false alarms that could overwhelm network managers, these metrics aid in evaluating the efficacy of the ML models. To improve the network's overall security posture, the suggested solution is integrated with the firewalls and intrusion prevention systems [IPS] that are already part of the network security infrastructure. The integration guarantees that automated defenses, such blocking harmful IP addresses or sending out alarm notifications, can be activated using the insights from big data analytics.

3.1 Problem Definition

i] Problem Statement
ii] Research Objectives
iii]Data Acquisition Tools
iv]Data Sources
v]Data Preparation
vi] Data Preprocessing includes data Cleansing, feature extraction, feature selection, data transformation
vii] Model Training & Testing Parameters:
Train and evaluate the models using the pre-processed features from the dataset.

3.2 Big Data Analytics Techniques:

Anomaly Detection: Implement statistical-based anomaly detection, clustering, or ML techniques. e.g.: RF [Random Forest etc.]
ML for Classification: Apply supervised learning algorithms [e.g., RF, DT] for classification. Train models on labeled datasets of security events [e.g., benign and malicious files].
Deep Learning: Use neural networks [e.g., CNNs or RNNs]. Implement advanced techniques like deep reinforcement learning for dynamic threat mitigation. Use graph-based models to detect suspicious patterns, i.e., relationships or potential attack vectors across network nodes.

3.3 Model Evaluation:

Scalability and Effectiveness: Evaluate the scalability of the analytics solution in terms of the volume of data processed, and the time taken to make decisions.

3.4 Implementation and Experimentation:

Deployment of Analytics Tools: Implement big data frameworks like Hadoop, Apache Spark, in link for processing and analyzing large datasets.
Simulation and Testing: Test the system under different network scenarios [e.g., botnet attacks, insider threats, DDoS attacks, etc.]. Simulate network traffic to assess the effectiveness of your system in detecting and mitigating threats.

3.5 Security and Privacy Considerations:

Data Privacy: Ensure that the data used for analysis complies with privacy standards such as GDPR or HIPAA.
Data Security: Employ encryption, access control, and secure data storage to protect sensitive information.
Threats to the Analytics System: Consider the possibility of attackers compromising the analytics system itself.

3.6 Results Analysis and Discussion:

Present the results of your model's performance in terms of detection accuracy, response time, and overall system effectiveness. Compare your approach with traditional network security methods [e.g., signature-based detection, rule-based systems]. Discuss the trade-offs, such as false positive rates, detection time, and the computational cost of big data analytics.

4. PROPOSED SYSTEM METHODOLOGY

The objective of the proposed System is to develop an advanced cyber-attack detection framework that integrates machine learning models with big data analytics to enhance security. This system will implement classification algorithms to distinguish between normal and malicious activity. Also detect various types of cyberattacks through automated analysis. The proposed methodology follows a structured flows for efficient cyberattacks detection. The next section illustrates the step process in a detailed flowchart, outlining key stages from data selection to attack detection and testing. We aim to leverage advanced analytics and ML techniques to detect and prevent cyber security threats. The system utilizes large-scale network traffic data

Figure 4: Flowchart of Explaining step by step proposed system methodology

5. DISCUSSION AND ANALYSIS

Ad hoc and novelty challenges are growing, as are labor-intensive data integration initiatives. This initial data and volume of data available to decision-makers is a significant obstacle to the entire process of existing and new scalable and semi-automated data integration solutions [23]. Ultimately, this study seeks to illuminate the path toward a more secure digital future. Where data-driven insights empower the ability to stay one step ahead of cybercriminals. Big data analytics enables monitoring and analyzing massive volumes of network traffic in near real-time, which may facilitate predictive capabilities, enabling proactive defenses rather than reactive measures. Given these factors, there is a pressing need to explore how big data frameworks can enhance network security. This research aims to develop a robust framework, underpinned by big data analytics, that not only enhances threat detection rates, but also minimizes the gap between big data analytics and cyber security practices and contribute valuable insights that empower organizations to safeguard their networks against the ever-evolving landscape of cyber threats.

The proposed framework successfully classified a variety of cyber-attacks, including phishing, malware, and brute-force attacks, showcasing its scalability and adaptability in dynamic network environments. To sum up, big data analytics is a significant paradigm shift in cybersecurity that provides a proactive, data-driven protection against changing threats. As cyberattacks grow in sophistication, further advancements in deep learning, real-time analytics, and edge computing will be crucial in fortifying cyber security. Organizations must embrace these innovations to safeguard critical infrastructure, ensuring a resilient and secure digital future.

REFERENCES

[1] A. Hussain, A. Memon and A. K. Memon, "Big Data for Network Security: A Survey", Int. Journal Computer Appl., Vol. 119, no. 4, PP. 19-24, 2015

[2] M. Ahmed, A. N. Mahmood, and J. Hu, "A Survey of Intrusion Detection System Using Machine Learning Techniques", journal Network Computer Appl. Vol.60, PP. 72-91, 2026

[3] Moustafa, N and Slay,J., "The Evaluation of Network Traffic Analysis Techniques for Intrusion Detection System", Int. Journal Computer Science Inf. Security, Vol.14, no.12, PP. 28-35, 2016

[4] Alazab M, Islam R., Khan A.," A Big Data Analytics for Intrusion Detection Systems: A Survey", Int. Journal Computer Science Inf. Security, Vol.15, no.9, PP. 68-75, 2017

[5] Jalali M., Sadegi M., Heidaril R., "A Big Data Analytics Approach for Intrusion Detection System", Int. Journal Computer Science Inf. Security, Vol.16, no.4, PP. 123-130, 2018
[6] Chen L., Zhang J. and Yang H., "A Big Data Analytics for Network Security: A Survey" Int. Journal Computer Science Inf. Security, Vol.16, no.8, PP. 12-18, 2018
[7] Zang X., Zhang Z., and Wang S., "A Big Data- Driven Intrusion Detection System: A Survey and future Directions", Int. Journal Computer Science Inf. Security, Vol.17, no.2, PP. 45-52, 2019
[8] F. Khan, L. Alazab and A. Zomaya, "A Big Data Analytics for APT Detection in Network Security", Computer Security, vol.87, PP101570,2019
[9] Y. Liu, H. Wang, and Z. Zhao, "Big Data- Driven intrusion Detection System in Network Security", Future Generation Computer System, vol.102, PP.223-235. 2020.
[10] Zhou Y., Li. X, & Zhang L.," Big Data Analytics for Network Security Techniques, Applications and Future Challenges", Int. Journal Computer Science Inf. Security, Vol. 18, no.6, PP 40-48, 2020
[11] Khan M., Jaman F. and Alazab M., "Big Data Analytics for Intrusion Detection and Prevention Systems", A Review Int. Journal Computer Science Inf. Security, Vol. 11, no.4, PP 13-19, 2020
[12] Kaur A., Singh M., & Sharma D.," Big Data Analytics for Network Security: Challenges and Solutions", Int. Journal Computer Science Inf. Security, Vol. 18, no.12, PP 35-42, 2020
[13] Deng L., Zhang W., & Liu Y.," Big Data Analytics for Network Security: A Survey and Future Research Directions", Int. Journal Computer Science Inf. Security, Vol. 19, no.5, PP 56-63, 2021
[14] Jiang Z, Chen X., & Liu H.," A Big Data Analytics for Intrusion Detection Systems: Techniques and Applications", Int. Journal Computer Science Inf. Security, Vol. 19, no.12, PP 89-96, 2021
[15] Wang X, Zhang Y. & Li. J., "Big Data Analytics in Network Security: A Revies of Techniques, Applications and Future Challenges" Int. Journal Computer Science Inf. Security, Vol. 19, no.8, PP 73-80, 2021
[16] Kaur A., Singh M., & Sharma D." Big Data Analytics for Network Security: Challenges and Solutions", Int. Journal Computer Science Inf. Security, Vol. 19, no.3, PP 45-52, 2021
[17] Bashir A., Khan M. S. & Ahmed M.," A Comprehensive review on Big Data Analytics for Enhancing Network Security", Int. Journal Computer Science Inf. Security, Vol. 20, no.7, PP 34-41, 2022
[18] Hussain A., Tariq U., & Iqbal M., "Big Data Analytics for Network Security: A Survey on Techniques and Applications", Int. Journal Computer Science Inf. Security, Vol. 20, no.5, PP 65-71, 2022
[19] Bhosale S., Patil A., & Kulkarni S., "Big Data Analytics for Intrusion Detection System: A Comprehensive Review and Future Directions", Int. Journal Computer Science Inf. Security, Vol. 20, no.9, PP 22-30, 2022
[20] Hussan M., Malik M. H., & Abbas F., "Advancement in Big Data Analytics for Enhancing Network Security: Challenges and Future Perspectives", Int. Journal Computer Science Inf. Security, Vol. 21, no.3, PP 45-52, 2023
[21] Cheng X., Li. Y & Zhang J.," Big Data Analytics for Network Security: Recent Advances & Future Challenges", Int. Journal Computer Science Inf. Security, Vol. 21, no.5, PP 12-18, 2023
[22] Kim S., Lee C., & Park H., "Anomaly Detection in Network Security Using Big Data & Deep Learning", Int. Journal Computer Science Inf. Security, Vol. 22, no.9, PP 94-101, 2023
[23] Costa A., Souza R., Silva T., "Enhancing Network Intrusion Detection Systems with Big Data and Machine Learning", Int. Journal Computer Science Inf. Security, Vol. 22, no.10, PP 102-109, 2023
[24] William T., Green D., & Robert P., "Big Data Analytics for Cyber Security: Emerging Trends and Techniques", Int. Journal Computer Science Inf. Security, Vol. 22, no.1, PP 12-19, 2023
[25] Taylor R., Smith J., Edward S., "Leveraging Big Data for Advanced Network Security and Monitoring", Int. Journal Computer Science Inf. Security, Vol. 22, no.2, PP 30-38, 2023